On shrinkage estimation for balanced loss functions [1]

Éric Marchand[a] & William E. Strawderman[b]
a Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1 (e-mail: eric.marchand@usherbrooke.ca)
b Rutgers University, Department of Statistics, 501 Hill Center, Busch Campus, Piscataway, N.J., USA, 08855 (e-mail: straw@stat.rutgers.edu)

Summary

The estimation of a multivariate mean θ is considered under natural modifications of balanced loss function of the form: (i) $\omega\,\rho(\|\delta-\delta_0\|^2)+(1-\omega)\,\rho(\|\delta-\theta\|^2)$, and (ii) $\ell\,(\omega\,\|\delta-\delta_0\|^2+(1-\omega)\,\|\delta-\theta\|^2)$, where δ_0 is a target estimator of $\gamma(\theta)$. After briefly reviewing known results for original balanced loss with identity ρ or ℓ, we provide, for increasing and concave ρ and ℓ which also satisfy a completely monotone property, Baranchik-type estimators of θ which dominate the benchmark $\delta_0(X)=X$ for X either distributed as multivariate normal or as a scale mixture of normals. Implications are given with respect to model robustness and simultaneous dominance with respect to either ρ or ℓ.

AMS 2010 subject classifications: 62F10, 62J07 (primary); 62C15, 62C20 (secondary).

Keywords and phrases: Balanced loss; Concave loss; Dominance; Multivariate normal; Scale mixture of normals; Shrinkage estimation.

1. Introduction

Balanced loss functions and their role in estimation have captured the interest of many researchers over the years since Arnold Zellner (Zellner, 1994) proposed their use in a regression framework. Balanced loss functions are appealing as they combine proximity of a given estimator δ to both a target estimator δ_0 and the unknown parameter θ which is being estimated. They relate conceptually to methods for combining estimators (e.g., Judge & Mittlehammer, 2004), as well as penalized least-squares estimation. The study of balanced loss functions has frequently been cast in a regression framework (e.g., Hu & Peng, 2011, and the references therein), but it also has arisen or related to credibility theory, finance, sequential estimation, etc (Baran & Stepień-Baran, 2013; Zhang & Chen, 2018). In Zellner's framework, the target estimator was least-squares, but such a target can be viewed more broadly (e.g., Jafari Jozani et al., 2006, 2014).

To a large extent, findings in the literature relate to balanced squared error loss

$$L_\omega(\theta,\delta)\,=\,\omega\,\|\delta-\delta_0\|^2\,+\,(1-\omega)\,\|\delta-\gamma(\theta)\|^2\,, \tag{1.1}$$

where for an observable $X\sim f_\theta$, $\gamma(\theta)\in\Gamma\subset\mathbb{R}^d$, $\delta_0(X)$ is a target estimator of $\gamma(\theta)$, $\omega\in[0,1)$ is the weight given to the proximity of δ to δ_0, and $\delta(X)$ is a given estimator of $\gamma(\theta)$. In such cases, as presented by Jafari Jozani et al. (2006), as well as Dey et al. (1999), Bayesian estimation as well as

[1] March 29, 2019

the frequentist risk performance under balanced loss L_ω with $\omega > 0$ relate precisely to corresponding features under unbalanced loss (i.e., squared error loss) L_0 (see Theorem 2.2). For instance, given a prior π and a corresponding Bayes estimator $\delta_{\pi,0}(X)$ under loss L_0, the corresponding Bayes estimator under balanced loss L_ω is simply given by $(1-\omega)\,\delta_\pi(X) + \omega\delta_0(X)$. Such relationships are reviewed and briefly illustrated in Section 2.

In contrast, much less is known for the following two natural alternatives or modifications to loss (1.1):

$$\omega\,\rho(\|\delta - \delta_0\|^2) + (1-\omega)\,\rho(\|\delta - \gamma(\theta)\|^2) \tag{1.2}$$

and

$$\ell\left(\omega\,\|\delta - \delta_0\|^2 + (1-\omega)\,\|\delta - \gamma(\theta)\|^2\right)\,, \tag{1.3}$$

with $0 \le \omega < 1$, $\rho(\cdot) \ge 0$, and $\ell(\cdot) \ge 0$. Balanced loss functions of the type (1.2) were considered by Jafari et al. (2012). They provided Bayesian estimators as well as other type of posterior risk analysis. However, for both losses (1.2) and (1.3), there seems to be no significant known finding for frequentist risk analysis, such as the earlier results for balanced squared error loss.

The objective of this paper is to try to fill such gaps. To achieve this, we focus on the multivariate normal case $X \sim N_d(\theta, \sigma^2 I_d)$, as well as scale mixture of normals as defined in (2.4), the target estimator $\delta_0(X) = X$; and the objective of improving on X. The latter is the maximum likelihood estimator and also is minimax for losses (1.2) as elaborated upon at the outset of Section 3C. We obtain various sufficient conditions for dominance for both losses (1.2) and (1.3). These apply for interesting subclasses of concave ρ's and ℓ's respectively, which are also completely monotone. Shrinkage estimation for multivariate normal models, and more generally spherically symmetric and elliptically symmetric models, has had a long, rich and influential history (e.g., Fourdrinier et al., 2018). The use of a concave loss as well as concave versions of (1.2) and (1.3), is quite appealing, and has motivated previous shrinkage estimation work such as Brandwein & Strawderman (1980, 1991), Brandwein et al. (1993), and Kubokawa et al. (2015), among others.

The paper is organized as follows. We collect some preliminary definitions and results in Section 2.1, before reviewing and illustrating frequentist risk and Bayesian analysis results in Section 2.2 applicable to balanced squared-error loss L_ω. In Sections 3 and 4, we provide conditions for a Baranchik-type estimator to dominate $\delta_0(X) = X$ under loss functions (1.2) and (1.3) respectively (i.e., Theorems 3.3 and 4.4). In both cases, the proofs are unified with respect to choice of model and loss, the former with respect to the underlying normal mixture and the latter with respect to the choice of ρ or ℓ for the balanced loss. Implications are given in terms of robustness and simultaneous dominance (i.e., Corollary 4.3). Finally, we make use of various techniques and properties relative to concave functions, completely monotone functions, superharmonic functions, and spherically symmetric distributions.

2. Preliminary results and the balanced squared-error loss case

2.1. Preliminary definitions and properties

We assemble here some definitions and properties useful throughout the manuscript. The estimators studied below are based on spherically symmetric distributions $X \sim f(\|x-\theta\|^2), x, \theta \in \mathbb{R}^d$, which are scale mixtures of normals. Such distributions admit the representation

$$X|V \sim N_d(\theta, VI_d),\ V \sim g\,, \tag{2.4}$$

and include many familiar examples such as Normal, Student, Logistic, Laplace, Exponential power (with $f(t) = t^s, 0 < s < 1$), among others. Other than moment finiteness conditions and the restriction to $d \geq 3$ or $d \geq 4$ dimensions, the applicability of our dominance findings will not require any further specific assumptions on f.

A key characterization and property, which brings into play completely monotone functions, is given by the following result (see, e.g., Feller, 1966; Berger, 1975; etc.).

Lemma 2.1. **(a)** *A density of the form $f(\|x-\theta\|^2), x, \theta \in \mathbb{R}^d$ is a scale mixture of normals if and only if $f(\cdot)$ is completely monotone, i.e., $(-1)^n f^n(t) \geq 0$, for $n = 1, 2, \ldots$, and $t \in \mathbb{R}_+$.*

(b) *The product of two completely monotone functions is completely monotone.*

The dominance findings of Sections 3 and 4 relate to Baranchik-type estimators of θ defined and denoted throughout as:

$$\delta_{a,r(\cdot)}(X) = \left(1 - \frac{ar(\|X\|^2)}{\|X\|^2}\right) X\,, \tag{2.5}$$

with $a > 0$, and the conditions

$$0 \leq r(\cdot) \leq 1\,,\ r(\cdot) \neq 0\,,\ r'(\cdot) \geq 0\,,\ \text{ and } (d/dt)(r(t)/t) \leq 0\,. \tag{2.6}$$

These include James-Stein estimators with constant $r(\cdot)$, and $r(t) = (d-2)\,\sigma^2$ in the original $X \sim N_d(\theta, \sigma^2 I_d)$ case.

2.2. Balanced squared-error loss

We review here, for Bayesian inference and frequentist risk analysis, relationships between balanced loss L_ω and its unbalanced counterpart L_0. Such results appear in Dey et al. (1999), as well as in Jafari Jozani et al. (2006). For the former, the findings apply to a multivariate normal model $X \sim N_d(\theta, \sigma^2 I_d)$ and $\delta_0(X) = X$, while the latter work relates to a more general model $X \sim f_\theta$ and target estimator δ_0. Some of the results will serve in later sections, but they are exposed here also to illustrate the facility in which Bayesian analysis and frequentist risk evaluations for L_ω follow from corresponding results for squared-error loss L_0.

The following Lemma 2.2 will be used in Section 4 for the analysis of losses in (1.3), but is presented here as it serves to link the frequentist risk under loss L_ω to the risk under squared error loss L_0, as presented in Corollary 2.1. To facilitate the presentation that follows, we denote the difference in losses L_ω between estimates $\delta_0(x) + (1-\omega)\, g(x)$ and $\delta_0(x)$ as

$$\Delta_\omega(\theta, g) \;=\; L_\omega\,(\theta, \delta_0 + (1-\omega) g) \;-\; L_\omega\,(\theta, \delta_0\,)\,. \tag{2.7}$$

Lemma 2.2. *Let $X \sim f_\theta$. For the problem of estimating $\gamma(\theta)$ under balanced loss L_ω (as in (1.1)), we have $\Delta_\omega(\theta, g) \;=\; (1-\omega)^2\, \Delta_0(\theta, g)$.*

Proof. A decomposition of (2.4) yields

$$\begin{aligned} \Delta_\omega(\theta, g) &= \omega\,\|\delta_0 + (1-\omega)g - \delta_0\|^2 \;+\; (1-\omega)\,\|\delta_0 + (1-\omega)g - \gamma(\theta)\|^2 \;-\; (1-\omega)\|\delta_0 - \gamma(\theta)\|^2 \\ &= (1-\omega)^2\,\|g\|^2 \;+\; 2(1-\omega)^2 g^\top(\delta_0 - \gamma(\theta)) \\ &= (1-\omega)^2\,\left(\|\delta_0 + g - \gamma(\theta)\|^2 - \;\|\delta_0 - \gamma(\theta)\|^2\right) \\ &= (1-\omega)^2\,\Delta_0(\theta, g). \qquad \square \end{aligned}$$

In terms of the frequentist risk R_ω associated with loss L_ω, given by $R_\omega(\theta, \delta) \;=\; \mathbb{E}\,\{L_\omega\,(\theta, \delta(X))\,\}$ for an estimator $\delta(X)$ of $\gamma(\theta)$, the following general result follows from Lemma 2.2.

Corollary 2.1. *Let $X \sim f_\theta$ and consider the problem of estimating $\gamma(\theta)$. The estimator $\delta_{1,\omega}(X) = \delta_0(X) + (1-\omega)g_1(X)$ dominates $\delta_{2,\omega}(X) \;=\; \delta_0(X) + (1-\omega)g_2(X)$ under loss L_ω if and only if $\delta_{1,0}(X) \;=\; \delta_0(X) + g_1(X)$ dominates $\delta_{2,0}(X) \;=\; \delta_0(X) + g_2(X)$ under squared error loss L_0.*

Proof. We have

$$\begin{aligned} R_\omega(\theta, \delta_{1,\omega}) \;-\; R_\omega(\theta, \delta_{2,\omega}) &= \mathbb{E}_\theta\{\Delta_\omega(\theta, g_1(X)) - \Delta_\omega(\theta, g_2(X))\} \\ &= (1-\omega)^2\;\mathbb{E}_\theta\{\Delta_0(\theta, g_1(X)) - \Delta_0(\theta, g_2(X))\} \quad \text{(Lemma 2.2)} \\ &= (1-\omega)^2\;\{R_0(\theta, \delta_{1,0}) \;-\; R_0(\theta, \delta_{2,0})\}, \end{aligned}$$

which establishes the result. $\square$

Now, turning to Bayesian inference, we have an equally simple relationship between balanced loss L_ω and its unbalanced counterpart L_0. More precisely, the following well-known result conveniently expresses the Bayes estimator $\delta_{\pi,\omega}$ under L_ω for $\omega > 0$ in terms of the Bayes estimator $\delta_{\pi,0}$ under L_0, given of course by $\delta_{\pi,0}(X) \;=\; \mathbb{E}(\gamma(\theta)|X)$.

Theorem 2.1. *For $X \sim f_\theta$ and a prior $\theta \sim \pi$ for which $Cov(\gamma(\theta)|x)$ exists for all x, the Bayes estimator $\delta_{\pi,\omega}$ of $\gamma(\theta)$ under loss L_ω is given by $\delta_{\pi,\omega}(X) \;=\; \omega\,\delta_0(X) \;+\; (1-\omega)\,\delta_{\pi,0}(X)\,$.*

Proof. Write $\delta_{\pi,\omega}(x) \;=\; \delta_0(x) \;+\; (1-\omega)\,g_{\pi,\omega}(x)$ for $0 \le \omega < 1$. By definition of the Bayes estimate, we thus have

$$\begin{aligned} g_{\pi,\omega}(x) &= \arg\min_g \mathbb{E}\{L_\omega\,(\theta, \delta_0(x) \;+\; (1-\omega)g)\;|\,x\} \\ &= \arg\min_g \mathbb{E}\{\Delta_\omega(\theta, g)\,|\,x\} \\ &= \arg\min_g \;\left[(1-\omega)^2\,\mathbb{E}\{\Delta_0(\theta, g)\,|\,x\}\right] \quad \text{(Lemma 2.2)} \\ &= \arg\min_g \mathbb{E}\{\|\delta_0(x) \;+\; g \;-\; \gamma(\theta)\|^2\, x\} \\ &= \mathbb{E}\{\gamma(\theta)|x\} \;-\; \delta_0(x)\,. \end{aligned}$$

From this, the result follows as $\delta_{\pi,\omega}(x) = \delta_0(x) + (1-\omega)\,[\mathbb{E}\{\gamma(\theta)|x\} - \delta_0(x)\,] = \omega\,\delta_0(x) + (1-\omega)\,\delta_{\pi,0}(x)\,.$ □

Now, combining the last two results leads to the following Bayes dominance result.

Corollary 2.2. *For $X \sim f_\theta$, a prior $\theta \sim \pi$, and the problem of estimating $\gamma(\theta)$, the Bayes estimator $\delta_{\pi,\omega}(X)$ dominates an estimator $\delta_0(X) + (1-\omega) g_2(X)$ under L_ω if and only if the Bayes estimator $\delta_{\pi,0}(X) = \mathbb{E}(\gamma(\theta)|X)$ dominates $\delta_0(X) + g_2(X)$ under squared-error loss L_0.*

Several examples can be found in the literature, namely among the references mentioned above. We do provide at the end of this subsection Example 2.1 as an illustration. Before doing so, we briefly address the issue of minimaxity, where relationships between balanced and unbalanced losses are not as immediate (e.g., Jafari Jozani et al., 2006). One situation though does simplify, namely the case where the target estimator $\delta_0(X)$ is itself minimax under the unbalanced loss. Moreover, the following result (Jafari Jozani et al., 2012, Theorem 4) holds in general for losses (1.2). As discussed at the outset of Section 3C, this will serve to guarantee that the dominating estimators of Theorem 3.3 are themselves minimax.

Theorem 2.2. *Let $X \sim f_\theta$ and consider the problem of estimating $\gamma(\theta)$ under loss (1.2). Suppose that the estimator $\delta_0(X)$ is minimax under unbalanced loss $\rho(\|\delta - \gamma(\theta)\|^2)$. Then, $\delta_0(X)$ is also minimax under loss (1.2) for all $0 < \omega < 1$.*

Proof. Let R_w denote the frequentist risk under loss (1.2). Since $R_\omega(\theta, \delta_0) = (1-\omega) R_0(\theta, \delta_0)$, the result is immediate. □

Example 2.1. *We consider the classical problem of estimating a multivariate normal mean and illustrate how known Stein estimation results applicable (e.g., Stein 1981; Strawderman, 2003) to squared-error loss L_0 translate to balanced loss L_ω. Let $X \sim N_d(\mu, \sigma^2 I_d)$ and $S^2 \sim \sigma^2 \chi^2_k$ be independently distributed with $d \geq 3, k \geq 1$. Set $\theta = (\mu, \sigma^2)$ and consider estimating $\tau(\theta) = \mu$ under balanced loss L_ω with target estimator $\delta_0(X) = X$, which is minimax under $\frac{L_0(\theta,\delta)}{\sigma^2}$, and thus minimax under loss $\frac{L_\omega(\theta,\delta)}{\sigma^2} = \frac{\omega\,\|\delta - X\|^2 + (1-\omega)\,\|\delta-\theta\|^2}{\sigma^2}$ by virtue of Theorem 2.2.*

- **(A)** *For known σ^2, any estimator of the form $\delta(X) = X + \sigma^2 g(X)$ such that g is weakly differentiable, $\mathbb{E}_\theta \|g(X)\|^2 < \infty$, and $\|g(X)\|^2 + 2\,div(g(X)) \leq 0$ a.e., dominates $\delta_0(X)$ under loss L_0. It thus follows from Corollary 2.1 that $X + (1-\omega)\,\sigma^2 g(X)$ dominates $\delta_0(X)$ under loss L_ω for such g's. Such dominating estimators include the James-Stein estimator with $g(t) = -(d-2)t/\|t\|^2$, as well as Baranchik type estimators $\delta_{a,r(\cdot)}(X)$ in (2.5), with $0 < a < 2(d-2)(1-\omega)\sigma^2$ and conditions (2.6) on $r(\cdot)$.*

- **(B)** *For unknown σ^2, with the same conditions on g, estimators of the form $X + \frac{S^2}{k+2}\, g(X)$ dominate X under loss L_0. Again, it follows immediately from Corollary 2.1 that estimators $X + (1-\omega)\,\frac{S^2}{k+2}\, g(X)$ dominate X under balanced loss L_ω.*

- **(C)** *For known σ^2, Bayes estimators $\delta_{\pi,\omega}(X)$ under balanced loss L_ω and associated with prior density $\pi(\theta)$, Theorem 2.1 along with a well-known representation for $\delta_{\pi,0}(X)$ tell us that*

$$\begin{aligned}\delta_{\pi,\omega}(X) &= \omega X + (1-\omega)\delta_{\pi,0}(X) \\ &= \omega X + (1-\omega)(X + \sigma^2 \frac{\nabla m(X)}{m(X)}) = X + (1-\omega)\,\sigma^2 \frac{\nabla m(X)}{m(X)}\,,\end{aligned}$$

where $m(X) = (2\pi\sigma^2)^{-1} \int_{\mathbb{R}^d} e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2} \pi(\theta)\, d\theta$ *is the marginal distribution of* X*. By virtue of Corollary 2.2, the estimator* $\delta_{\pi,\omega}(X)$ *dominates* X *under loss* L_ω *if and only if* $\delta_{\pi,0}(X)$ *dominates* X *under loss* L_0*. With the superharmonicity of either* $\pi(\cdot), m(\cdot)$ *or* $\sqrt{m(\cdot)}$ *a sufficient condition for* $\delta_{\pi,0}(X)$ *to dominate* X *under loss* L_0 *(e,g. Strawderman, 2003), we thus infer that either of these conditions imply that* $\delta_{\pi,\omega}(X)$ *dominates* X *under balanced loss* L_ω*.*

3. Risk analysis for loss $\omega\rho(\|\delta - X\|^2) + (1-\omega)\rho(\|\delta - \theta\|^2)$

A. The loss function

For a model (2.4), we evaluate the frequentist risk performance of an estimator $\delta(X)$ of θ under the balanced loss

$$L_{\omega,\rho}(\theta,\delta) \;=\; \omega\, \rho(\|\delta - X\|^2) + (1-\omega)\, \rho(\|\delta-\theta\|^2)\,, 0 \le \omega < 1\,, \tag{3.8}$$

which incorporates the target estimator $\delta_0(X) = X$. For the function ρ, we assume the following throughout this section:

Assumption 1. $\rho(0) = 0\,,\ 0 < \rho'(0) < \infty$, *and* ρ' *is completely monotone on* $\mathbb{R}_+$*, i.e.,* $(-1)^k\, \rho^{(k+1)}(t) \geq 0$ *for* $t > 0$ *and for* $k = 0, 1, \ldots$*.*

Examples of loss functions $L_{\omega,\rho}$ for which ρ satisfies Assumption 1, other than $\rho(t) = t$, include: **(i)** $\rho(t) = 1 - e^{-t/\alpha}$ with $\alpha > 0$, **(ii)** $\rho(t) = \log(1+t)$, **(iii)** $\rho(t) = (1 + t/\gamma)^\beta$ with $\gamma > 0, \beta \in (0,1)$, and **(iv)** cases $\rho(t) = z(0) - z(t)$ with z being completely monotone such as $\rho(t) = r^2 t/(rt+1)$ with $r > 0$. Case **(i)** is known as reflected normal loss, while examples **(iv)** represent a broader class of bounded losses. L^β losses with $\rho(t) = t^\beta$, $\beta \in (0,1)$, represent concave choices, but such ρ's do not satisfy the finiteness assumption on $\rho'(0)$.

B. Further technical results

We now expand on various technical results which are pivotal to the risk analysis in Subsection **3C**.

Lemma 3.3. *Consider* $X \sim f(\|x-\theta\|^2), x, \theta \in \mathbb{R}^d$*, admitting representation (2.4) with mixing variable* V*, and* ρ *satisfying Assumption 1. Let* $Y \sim f^*(\|y-\theta\|^2) \;=\; \rho'(\|y-\theta\|^2)\, f(\|y-\theta\|^2)/K\,; y \in \mathbb{R}^d$*; with* $K = \mathbb{E}_0(\rho'(\|X\|^2))$*. Then,*

(a) *The distribution of* Y *admits a scale mixture of normals representation*

$$Y|W \sim N_d(\theta, W I_d)\,, \quad \text{with } W \sim h\,; \tag{3.9}$$

(b) *Moreover, the distribution of* W *is stochastically smaller than the distribution of* V*.*

Proof. First observe that f^* is a density since $K = \mathbb{E}_0(\rho'(\|X\|^2)) \leq \rho'(0) < \infty$. Part **(a)** thus follows from Lemma 2.1. For part **(b)**, given that ρ' and f are completely monotone, they are representable as Laplace transforms (Lemma 2.1):

$$f(t) = K_1 \int_0^\infty e^{-t/2v}\, dG(v)\,,$$
$$\rho'(t) = K_2 \int_0^\infty e^{-t/2\tau}\, dH(\tau)\,,$$

for $t \in \mathbb{R}^d$. From this, we have for $t \in \mathbb{R}^d$

$$\begin{aligned} f^*(t) &= \frac{K_1 K_2}{K} \int_0^\infty \int_0^\infty e^{-\frac{t}{2}(\frac{1}{v}+\frac{1}{\tau})}\, dG(v)\, dH(\tau)\,, \\ &= K_3 \int_0^\infty e^{-t/2w} dM(w)\,. \end{aligned}$$

Interpreting in terms of scale mixture of normals, we have for $Y \sim f^*(\|y-\theta\|^2)$ representation (3.9) with $W =^d \frac{\tau V}{\tau + V}$. Finally, from this, we have $\mathbb{P}(W \leq s) \geq \mathbb{P}(V \leq s)$, for all $s > 0$ and the result follows. □

The two lemmas that follow, which we will require, rely partly on properties of superharmonic functions. We recall that a continuous function $g : \mathbb{R}^d \to \mathbb{R}$ is superharmonic if and only if: at all $t_0 \in \mathbb{R}^d$ and $r > 0$, the average of g over the surface of the sphere, centered at t_0 of radius r, $S_r(t_0) = \{t \in \mathbb{R}^d : \|t - t_0\| = r\}$ is less or equal than $g(t_0)$. For twice differentiable g, the superharmonicity of g is equivalent to its Laplacian being less or equal to 0, i.e., $\Delta\, g \leq 0$ with $\Delta\, g = \sum_{i=1}^d \frac{\partial^2}{\partial t_i^2} g(t)$.

Lemma 3.4. *Let $Z \sim N_d(0, I_d)$ with $d \geq 3$ and let $T = \|\alpha Z + \theta\|^2$ with $\alpha > 0$ and $\theta \in \mathbb{R}^d$. Then, we have the following:*

- **(a)** *$\epsilon(\alpha) = \mathbb{E}_\alpha(\frac{1}{T})$ is decreasing in α for $\alpha > 0$;*
- **(b)** *$\mathbb{E}\, g(\alpha Z + \theta)$ is non-increasing in α provided that $g(\cdot) > 0$ and g is superharmonic.*

Proof. The proof of part **(a)** is relegated to an Appendix. For part **(b)**, first denote $U_m, m > 0$, as a random vector uniformly distributed on the sphere $S_m(0)$ centered at 0 of radius m. It suffices to show that $\beta(\alpha, r) = \mathbb{E}\ (g(\alpha Z + \theta) | \|Z\| = r)$ is for all $r > 0$ decreasing in α. Since $(Z | \|Z\| = r) \sim U_r$ independently of $\|Z\|$, we have $\beta(\alpha, r) = \mathbb{E}\ \{g(U_{\alpha r} + \theta)\}$. Since, for a superharmonic function, the sphere mean is decreasing in the radius (see, e.g., Fourdrinier et al. 2018, Theorem 7.4), we infer that $\beta(\alpha, r)$ is decreasing in α, which concludes the proof. □

Lemma 3.5. *Let $\theta \in \mathbb{R}^d$, $a > 0$, and ρ satisfy Assumption 1. Consider $X \sim f(\|x - \theta\|^2)$, $Y \sim f^*(\|y - \theta\|^2)$ as in (2.4) and Lemma 3.3, respectively.*

- **(a)** *For $d \geq 3$, we have*

$$\mathbb{E}_\theta \left(\rho(\frac{a^2}{X'X}) \right) \leq \rho'(0)\, \mathbb{E}_\theta(\frac{a^2}{X'X}) \leq \rho'(0)\, \mathbb{E}_\theta(\frac{a^2}{Y'Y})\,;$$

(b) *For $d \geq 4$ and $r : \mathbb{R}^d \to [0,1]$ a twice-differentiable function that is non-decreasing and concave, we have*

$$\mathbb{E}_\theta \left(\rho(\frac{a^2 \, r^2(\|X\|^2)}{\|X\|^2}) \right) \leq \rho'(0) \, \mathbb{E}_\theta \left(\frac{a^2 \, r(\|Y\|^2)}{\|Y\|^2} \right) .$$

Proof. **(a)** The first inequality follows from the inequality

$$\rho(t) \leq \rho(0) + \rho'(0)t = \rho'(0) \, t \,, \tag{3.10}$$

which holds since ρ is concave with $\rho(0) = 0$. The second inequality follows from Lemma 3.3 and part (a) of Lemma 3.4. Indeed, since $X|V \sim N_d(\theta, VI_d)$, $Y|W \sim N_d(\theta, WI_d)$, we have, with the notation of Lemma 3.4, $\mathbb{E}_\theta(\frac{1}{\|X\|^2}) = \mathbb{E}(\epsilon(\sqrt{V}))$ and $\mathbb{E}_\theta(\frac{1}{\|Y\|^2}) = \mathbb{E}(\epsilon(\sqrt{W}))$, and the result follows since $\epsilon(\cdot)$ is decreasing and W is stochastically smaller than V.

(b) Defining $Z \sim N_d(0, I_d)$ and denoting $g_0(t) = \frac{r(\|t\|^2)}{\|t\|^2}$, $t \in \mathbb{R}^d$, we have

$$\begin{aligned} \mathbb{E}_\theta \left(\rho(\frac{a^2 \, r^2(\|X\|^2)}{\|X\|^2}) \right) &\leq \mathbb{E}_\theta \left(\rho(\frac{a^2 \, r(\|Y\|^2)}{\|Y\|^2}) \right) \\ &\leq a^2 \, \rho'(0) \, \mathbb{E}_\theta \left(\frac{r(\|X\|^2)}{\|X\|^2} \right) \\ &= a^2 \, \rho'(0) \, \mathbb{E} \left(g_0(\sqrt{V} \, Z + \theta) \right) \\ &\leq a^2 \, \rho'(0) \, \mathbb{E} \left(g_0(\sqrt{W} \, Z + \theta) \right) \\ &= a^2 \, \rho'(0) \, \mathbb{E}_\theta \left(\frac{r(\|Y\|^2)}{\|Y\|^2} \right) , \end{aligned}$$

where (i) the two equalities follow from the scale mixture representations of f and f^*; (ii) the first inequality follows since ρ is non-decreasing and $0 \leq r^2(\|t\|^2) \leq r(\|t\|^2) \leq 1$ for $t \in \mathbb{R}^d$, (iii) the second inequality follows from (3.10), and (iv) the third inequality follows from Lemma 3.3, part **(b)** of Lemma 3.4, as in the above proof of part **(a)**, and from the fact that

$$g_0(t) \text{ is superharmonic for } d \geq 4, \tag{3.11}$$

provided $r(\cdot)$ is non-negative, non-decreasing, and concave. Finally, to justify the above, note that, for twice-differentiable $h(\|t\|^2), t \in \mathbb{R}^d$,

$$\Delta \, h(\|t\|^2) = 2dh'(\|t\|^2) + (t't) \, h''(\|t\|^2) \,,$$

so that the choice $h(\|t\|^2) = g_0(t) = \frac{r(\|t\|^2)}{\|t\|^2}$ yields with a little bit of computation

$$\begin{aligned} \Delta \, (g_0(t)) &= 2 \left((\frac{d-4}{\|t\|^2}) \, \{\|t\|^2 \, r'(\|t\|^2) - r(\|t\|^2)\} + 2r''(\|t\|^2) \right) \\ &\leq 0 \end{aligned}$$

since the properties of $r(\cdot)$ imply that $r''(u) \leq 0$ and $r(u) \geq ur'(u)$ for all $u > 0$. □

C. Dominance results

For balanced loss $L_{\omega,\rho}$ with ρ satisfying Assumption 1, a scale mixture of normals distribution on X with $d \geq 3$, we provide James-Stein and Baranchick-type estimators that dominate X. In such cases, it follows that X is minimax for the unbalanced case $L_{0,\rho}$ with constant risk R_0 (e.g., Kubokawa et al., 2015). By virtue of Theorem 2.2, X is also minimax for balanced loss $L_{\omega,\rho}$. The following dominance results thus provide dominating estimators which are also minimax under loss $L_{\omega,\rho}$.

Theorem 3.3. *Consider $X \sim f(\|x-\theta\|^2)$; $x, \theta \in \mathbb{R}^d$; admitting representation (2.4), balanced loss function $L_{\omega,\rho}$ as in* (3.8) *with ρ satisfying Assumption 1.*

(a) *If $d \geq 3$, $\delta_a(X) = (1 - \frac{a}{\|X\|^2})X$ dominates $\delta_0(X) = X$ provided*

$$0 < a < \frac{2(d-2)K(1-\omega)\ \{\mathbb{E}(W^{-1})\}^{-1}}{\omega\rho'(0) + (1-\omega)K}\,, \tag{3.12}$$

with $K = \mathbb{E}_0(\rho'(\|X\|^2))$, and W the mixing variance for $Y \sim f^*(\|y-\theta\|^2)$ as defined in Lemma 3.3. An equivalent expression for the above dominance condition is

$$0 < a < \frac{2K^2(1-\omega)\ \left\{\mathbb{E}_0\left(\frac{\rho'(\|X\|^2)}{\|X\|^2}\right)\right\}^{-1}}{\omega\rho'(0) + (1-\omega)K}\,. \tag{3.13}$$

(b) *If $d \geq 4$, a Baranchik-type estimator $\delta_{a,r(\cdot)}(X)$ in (2.5) dominates $\delta_0(X) = X$ provided (3.12) holds and provided $r(\cdot)$ satisfies conditions (2.6).*

Proof. (a) First, the stated equivalence between (3.12) and (3.13) holds since, on one hand,

$$\mathbb{E}_0(\frac{1}{(\|Y\|^2)} = \mathbb{E}\{\frac{1}{W}\mathbb{E}_0(\frac{W}{\|Y\|^2}\,|W)\} = \frac{1}{d-2}\,\mathbb{E}\left(\frac{1}{W}\right)\,,$$

(as $\frac{\|Y\|^2}{W}|W \sim \chi^2_d(0)$ when $\theta = 0$) and, on the other hand,

$$\mathbb{E}_0(\frac{1}{(\|Y\|^2)} = \int_{\mathbb{R}^d} f^*(\|y\|^2)\,\frac{1}{(\|y\|^2)}\,dy = \frac{1}{K}\int_{\mathbb{R}^d} \rho'(\|x\|^2)\,f(\|x\|^2)\,\frac{1}{(\|x\|^2)}\,dx\,.$$

Second, we have for a difference in risks

$$\begin{aligned}
\Delta_a(\theta) &= R(\theta,\delta_a) - R(\theta,\delta_0)\\
&= \mathbb{E}_\theta\left[\omega\rho(\|\delta_a(X)-X\|^2) + (1-\omega)\,\rho(\|\delta_a(X)-\theta\|^2) - (1-\omega)\,\rho(\|X-\theta\|^2)\right]\\
&= \mathbb{E}_\theta\left[\omega\rho(\|\frac{aX}{\|X\|^2}\|^2) + (1-\omega)\left(\rho(\|(1-\frac{a}{\|X\|^2})X-\theta\|^2) - \rho(\|X-\theta\|^2)\right)\right]\\
&\leq \mathbb{E}_\theta\left[\omega\rho'(0)\,\frac{a^2}{\|Y\|^2} + (1-\omega)\,K\,\frac{\rho'(\|X-\theta\|^2)}{K}\left(\frac{a^2}{\|X\|^2} - \frac{2aX'(X-\theta)}{\|X\|^2}\right)\right]\\
&= \mathbb{E}_\theta\left[\omega\rho'(0)\,\frac{a^2}{\|Y\|^2} + (1-\omega)\,K\left(\frac{a^2}{\|Y\|^2} - \frac{2aY'(Y-\theta)}{\|Y\|^2}\right)\right],
\end{aligned}$$

where the inequality follows from part **(a)** of Lemma 3.5 and the concave function inequality $\rho(t_1)-\rho(t_2) \leq \rho'(t_1)(t_1-t_2)$ for all $t_1, t_2 \geq 0$. Now, with representation (3.9), by conditioning on W, and by the Stein's identity and calculation $\mathbb{E}_\theta\left[(Y-\theta)'\frac{Y}{Y'Y})|W\right] = W\,\mathbb{E}_\theta\,\mathrm{div}\left(\frac{Y}{Y'Y}\right) = W\,\mathbb{E}_\theta\left(\frac{d-2}{Y'Y}\right)$ (with probability 1), we obtain

$$\Delta_a(\theta) \;\leq\; a\,\mathbb{E}^W\left\{\mathbb{E}_\theta^{Y|W}\frac{W}{\|Y\|^2}\left(\frac{a\,(\omega\rho'(0)+(1-\omega)K)}{W} - 2(1-\omega)K(d-2)\right)\right\}. \tag{3.14}$$

By noticing that $\mathbb{E}_\theta^{Y|w}\left(\frac{W}{\|Y\|^2}\right)$ is increasing in $w>0$, given that $\frac{\|Y\|^2}{W}|W \sim \chi_d^2(\lambda = \frac{\|\theta\|^2}{W})$ and $\chi_d^2(\lambda)$ distributions are stochastically increasing in λ, we infer from (3.14) and the covariance inequality (i.e., $\mathbb{E}f_1(W)f_2(W) \leq \mathbb{E}f_1(W)\,\mathbb{E}f_2(W)$ for $f_1(\cdot)$ increasing and $f_2(\cdot)$ decreasing) that

$$\Delta_a(\theta) \;\leq\; a\,\mathbb{E}\left\{\mathbb{E}_\theta^{Y|W}\frac{W}{\|Y\|^2}\right\}\,\mathbb{E}\left(\frac{a\,(\omega\rho'(0)+(1-\omega)K)}{W} - 2(1-\omega)K(d-2)\right). \tag{3.15}$$

From the above, it follows immediately that (3.12) is a sufficient condition for $\Delta_a(\theta)$ to be negative for all θ.

(b) The proof is similar to that of part **(a)**. Using the concave function inequality $\rho(t_1)-\rho(t_2) \leq \rho'(t_1)(t_1-t_2)$, Stein's identity, and part **(a)** of Lemma 3.5, we obtain for the difference in risk

$$\begin{aligned}\Delta_a(\theta) &= R(\theta,\delta_{a,r(\cdot)}) - R(\theta,\delta_0)\\ &\leq a\,\mathbb{E}^W\left\{\mathbb{E}_\theta^{Y|W}\frac{W\,r(\|Y\|^2)}{\|Y\|^2}\left(\frac{a\,(\omega\rho'(0)+(1-\omega)K)}{W} - 2(1-\omega)K(d-2)\right)\right\}.\end{aligned}$$

Now, it is easy to verify that $r(t)/t$ is decreasing in $t>0$ under the given conditions on $r(\cdot)$. Finally, an application of the covariance inequality leads to an inequality as in (3.15) with $\mathbb{E}_\theta^{Y|W}\left(\frac{W}{\|Y\|^2}\right)$ replaced by $\mathbb{E}_\theta^{Y|W}\left(\frac{W\,r(\|Y\|^2)}{\|Y\|^2}\right)$. The result then follows. □

Remark 3.1. *From inequality (3.15), it also follows that dominance occurs, in both parts* **(a)** *and* **(b)** *of Theorem 3.3 for the quantity a equal to the upper cut-off point in (3.12) (or (3.13)) unless $\rho(t)=t$ and W is degenerate, i.e., original balanced loss and the multivariate normal case.*

The proof of Theorem 3.3 is unified with respect to the choice of ρ, the coefficient ω in the balanced loss, and the underlying scale mixture or normals distribution. To conclude, we point out that the above result can be seen as extensions of Kubokawa et al. (2015), as well as Strawderman (1974), whose results can be seen as particular cases of $\omega=0$ in the former case, and $\omega=0, \rho(t)=t$ in the latter case.

4. Risk analysis for loss $\ell\left(\omega\|\delta-\delta_0\|^2+(1-\omega)\|\delta-\theta\|^2\right)$

The main dominance finding of this section (Theorem 4.4) relates to a multivariate normal $X \sim N_d(\theta,\sigma^2 I_d)$, and more generally to $X \sim f(\|x-\theta\|^2)$ distributed as a scale mixture of normals as in (2.4). We assess the frequentist risk performance of an estimator $\delta(X)$ of $\gamma(\theta)$ under the balanced loss

$$L_{\omega,\ell}(\theta,\delta) \;=\; \ell\left(\omega\,\|\delta-\delta_0\|^2 + (1-\omega)\,\|\delta-\gamma(\theta)\|^2\right),\, 0 \leq \omega < 1, \tag{4.16}$$

More specifically, we consider the target estimator $\delta_0(X) = X$ and set $\gamma(\theta) = \theta$, and our objective is to provide, for $d \geq 3$, estimators of θ that dominate $\delta_0(X) = X$ under balanced loss (4.16) other than Section 2's results for $\ell(t) = t$. For the function ℓ, we assume, unless stated otherwise, the following throughout this section:

Assumption 2. *$\ell(\cdot) \geq 0$, $\ell'(\cdot) > 0$, and ℓ' is completely monotone on $\mathbb{R}_+$, i.e., $(-1)^k \, \ell^{(k+1)}(t) \geq 0$ for $t > 0$ and for $k = 0, 1, \ldots$.*

Examples of losses $L_{\omega,\ell}$ with ℓ satisfying Assumption 2 include examples **(i), (ii), (iii), (iv)** given for ρ in part **B.** of Section 3, but the cases $\ell(t) = t^\beta, \beta \in (0,1)$, are also included here since the assumption $\ell'(0) < \infty$ is not required.

We proceed with a preparatory lemma which exploits the concavity of ℓ, and which relates the difference in losses $L_{\omega,\ell}$, between estimates $\delta_0(x) + (1-\omega)\, g(x)$ and $\delta_0(x)$, to the balanced squared-error loss difference $\Delta_\omega(\theta, g)$ in (2.7). We therefore define

$$\Delta_{\omega,\ell}(\theta, g) \;=\; L_{\omega,\ell}\,(\theta, \delta_0 + (1-\omega) g) \;-\; L_{\omega,\ell}\,(\theta, \delta_0\,)\,, \tag{4.17}$$

and we have the following.

Lemma 4.6. *Let $X \sim f_\theta$. For the problem of estimating θ under loss (4.16) with twice-differentiable, increasing, and concave ℓ, we have*

$$\Delta_{\omega,\ell}\,(\theta, g) \;\leq\; (1-\omega)^2 \, \ell' \left\{ (1-\omega)\, \|\delta_0 - \gamma(\theta)\|^2 \right\} \, \Delta_0\,(\theta, g)\,. \tag{4.18}$$

Proof. The proof uses the fact that $\ell(a+b) - \ell(a) \leq b\, \ell'(a)$, since ℓ is concave, with $a = L_\omega(\theta, \delta_0)$ and $a + b = L_\omega(\theta, \delta_0 + (1-\omega) g)$. This yields :

$$\begin{aligned} \Delta_{\omega,\ell}(\theta, g) \;&=\; \ell\,(L_\omega(\theta, \delta_0 + (1-\omega) g)) \;-\; \ell\,(L_\omega(\theta, \delta_0)) \\ &\leq\; \ell'\,(L_\omega(\theta, \delta_0))\; \Delta_\omega(\theta, g)\,, \end{aligned}$$

which is indeed (4.18), by virtue of Lemma 2.2 and since $L_\omega(\theta, \delta_0) \;=\; (1-\omega)\, \|\delta_0 - \gamma(\theta)\|^2$. □

A basic result for estimating a mean vector θ under quadratic loss, for scale mixtures of normal distributions is the following.

Lemma 4.7. *(Strawderman, 1974) Let $X \sim f(\|x-\theta\|^2)$ have a scale mixture of normals distribution as in (2.4) with $d \geq 3$, and consider estimating θ with loss $\|\delta - \theta\|^2$. Consider Baranchik-type estimators $\delta_{a,r(\cdot)}(X)$ as in (2.5) with conditions (2.6). Then, $\delta_{a,r(\cdot)}(X)$ dominates X provided*

$$0 \,<\, a \,<\, \frac{2}{\mathbb{E}_0(\frac{1}{\|X\|^2})}\;,$$

and provided $E_0[\|X\|^2]$ and $E_0[1/||X||^2]$ are finite.

The main result of this section can now be presented and established.

Theorem 4.4. *Let $X \sim f(\|x-\theta\|^2)$ have a scale mixture of normals distribution as in (2.4) with $d \geq 3$, and consider estimating θ with loss $L_{\omega,\ell}$, as in (4.16) with $\delta_0(X) = X$, where ℓ' satisfies Assumption 2. Consider Baranchik-type estimator $\delta_{a(1-\omega),r(\cdot)}(X)$ as in (2.5) with conditions (2.6)* *on $r(\cdot)$. Then, assuming that $f_0^*(x) = \ell'((1-\omega)\|x\|^2)\, f(\|x\|^2)/K_1$ is a density on $\mathbb{R}^d$, $\delta_{a(1-\omega),r}(X)$ dominates X provided*

$$0 < a < \frac{2}{\mathbb{E}^*_{0,\omega}(\frac{1}{\|X\|^2})} \,,$$

and provided $E^*_{0,\omega}[\|X\|^2]$ and $E^*_{0,\omega}[1/\|X\|^2]$ are finite, where the expectation $\mathbb{E}^*_0$ is taken with respect to f_0^*.

Proof. With the given notation, observe that $\delta_{a(1-\omega),r}(X) = X + (1-\omega)\, g_{a,r}(X)$ with $g_{a,r}(X) = -\frac{ar(\|X\|^2)}{\|X\|^2}\, X$. Therefore, by Lemma 4.6 with $\delta_0(X) = X$ and $\gamma(\theta) = \theta$, we have for the difference in losses between $\delta_{a(1-\omega),r}(X)$ and X:

$$\begin{aligned}\Delta_{\omega,\ell}\,(\theta, g_{a,r}) &\leq (1-\omega)^2\, \ell'\left\{(1-\omega)\,\|X-\theta\|^2\right\}\, \Delta_0\,(\theta, g_{a,r}) \\ &= K_1\,(1-\omega)^2 \int_{\mathbb{R}^d} \left\{\|\delta_{a,r}(x)-\theta\|^2 \,-\, \|x-\theta\|^2\right\} f_0^*(\|x-\theta\|^2\, dx\,.\end{aligned}$$

Finally, since both ℓ' and f are completely monotone, so is the density f_0^* (Lemma 2.1, part a). This implies that f_0^* is a scale mixture of normals density and the result thus follows immediately from Lemma 4.7. □

Remark 4.2. *For the unbalanced case $\omega = 0$, one recovers Theorem 2.1 of Kubokawa et al. (2015). For the original balanced loss function with $\ell(t) = t$, one may recover the result of Theorem 4.4 directly by relying on Lemma 4.7 and Lemma 2.2, as illustrated in Example 2.1 for the multivariate normal case.*

Moreover, it is interesting to compare the balanced and unbalanced cut-off points $a_0(\omega) = \frac{2}{\mathbb{E}^*_{0,\omega}(\frac{1}{\|X\|^2})}$ and $a_0(0) = \frac{2}{\mathbb{E}^*_{0,0}(\frac{1}{\|X\|^2})}$. For loss $L_{\omega,\ell}$ with $\ell(t) = t^\beta$, $0 < \beta < 1$, we have $a_0(\omega) = a_0(0)$ for all $\omega \in (0,1)$. For choices ℓ such that $\frac{\ell'((1-\omega)t)}{\ell'(t)}$ is non-decreasing in t, we have a monotone likelihood ratio ordering for densities f_0^* and f, with the former being stochastically larger. This implies the ordering $\mathbb{E}^*_{0,\omega}(\frac{1}{\|X\|^2}) \leq \mathbb{E}^*_{0,0}(\frac{1}{\|X\|^2})$, and therefore $a_0(\omega) \geq a_0(0)$. Examples where such a condition holds include: $\ell(t) = 1 - e^{-t/\alpha}$, $\alpha > 0$, $\ell(t) = \log(1+t)$, $\ell(t) = (1 + t/\beta)^\alpha$ with $\beta > 0, 0 < \alpha < 1$.

It is also interesting to assess how the upper cut-off point on the multiple a for the estimator $\delta_{a,r(\cdot)}$ to dominate X varies in terms of the model f and the choice of ℓ for the loss function. In the former case, one can infer dominance results that are robust, holding for a given f but also persisting for a class of departures from f. This is quite plausible and simple to visualize as the cut-off point depends on f only through the inverse moment $\mathbb{E}^*_{0,\omega}(\frac{1}{\|X\|^2})$. For the latter case, one can infer dominance results that hold simultaneously for a subclass of losses (4.16). Here is such an illustration.

Corollary 4.3. *Consider the context of Theorem 4.4, for a given loss $L_{\omega,\ell}$, and a Baranchik-type estimator $\delta_{a(1-\omega),r(\cdot)}(X)$ which satisfies the given requirements for dominance of $\delta_0(X) = X$. Then, the dominance persists for the original balanced loss L_{w,ℓ_0} with $\ell_0(t) = t$ and $\delta_0(X) = X$.*

Proof. It suffices to show that

$$\mathbb{E}^*_0(\frac{1}{\|X\|^2}) \geq \mathbb{E}_0(\frac{1}{\|X\|^2})\,, \tag{4.19}$$

where the expectation $\mathbb{E}_0^*$ is taken with respect to the density f_0^* given in Theorem 4.4, and where we define the expectation $\mathbb{E}_0$ as taken with respect to the density $f(\|x\|^2)$. Now observe that the ratio of these densities, proportional to $\ell'((1-\omega)\|x\|^2)$ is decreasing in $\|x\|^2$ by assumptions on ℓ. We thus have a monotone likelihood ratio in $\|x\|^2$ ordering between the densities and inequality (4.19) follows since $1/\|x\|^2$ is decreasing in $\|x\|^2$. □

5. Concluding remarks

For a multivariate normal distributed $X \sim N_d(\theta, \sigma^2 I_d)$, and more generally for a scale mixture of normals model $X \sim f(\|x-\theta\|^2)$, we have provided shrinkage estimators of θ that improve on the benchmark estimator $\delta_0(X)$ as measured by the frequentist risk associated with balanced loss functions of the types (1.2) and (1.3), and with completely monotone ρ and ℓ. Much of the approach is unified with respect to the choices of f and either ρ or ℓ and the findings represent analytical extensions to the original balanced loss with either identity ρ or ℓ, unavailable up to now.

The findings in this paper do not cover cases with unknown scale such as observations generated from a $N_d(\theta, \sigma^2 I_d)$ with unknown σ^2, such as earlier results on the original balanced loss function (e.g., Chung et al. 1999; Zinodiny, 2014), but we expect that the techniques presented here should be useful to derive corresponding results for analogs of loss functions (1.2) and (1.3). Finally, it would be most interesting and welcomed to obtain Bayesian estimators that either satisfy our conditions of dominance, or dominate the benchmark $\delta_0(X) = X$ under the set-up of Theorems 3.3 and 4.4.

Appendix

Proof of Lemma 3.4, part (a)

With $T/\alpha^2 = \|Z + \frac{\theta}{\alpha}\|^2 \sim \chi_d^2(\|\theta\|^2/\alpha^2)$ and the Poisson representation of the non-central χ^2 distribution (i.e., $T/\alpha^2|K \sim \chi^2_{d+2K}$, $K \sim \text{Poisson}(\lambda = \|\theta\|^2/2\alpha^2)$), we have for $d \geq 3$ and $\theta \neq 0$

$$\begin{aligned}
\mathbb{E}_\alpha(\frac{1}{T}) &= \frac{1}{\alpha^2}\, e^{-\lambda} \sum_{k\geq 0} \frac{\lambda^k}{k!}\, \mathbb{E}[\frac{1}{\chi^2_{d+2k}}] \\
&= \frac{1}{\alpha^2}\, e^{-\lambda} \sum_{k\geq 0} \frac{\lambda^k}{k!}\, \frac{1}{d+2k-2} \\
&= \frac{2}{\|\theta\|^2}\, e^{-\lambda} \sum_{k\geq 0} \frac{\lambda^{k+1}}{(k+1)!}\, \frac{k+1}{d+2k-2} \\
&= \frac{2}{\|\theta\|^2}\, e^{-\lambda} \sum_{k\geq 0} \frac{\lambda^k}{k!}\, U(k) \\
&= \frac{2}{\|\theta\|^2}\, \mathbb{E}_\lambda\, U(K)\,, \qquad (5.20)
\end{aligned}$$

with $U(K) = \frac{K}{d+2K-4}\, \mathbb{I}_{\mathbb{N}_+}(K)$. Since $U(k)$ is increasing in $k \in \mathbb{N}$ for $d \geq 3$, since λ is decreasing in α, and since the Poisson(λ) distribution has increasing monotone likelihood ratio in K with parameter λ, it follows from the above that $\mathbb{E}_\alpha(\frac{1}{T})$ is decreasing indeed in α for $d \geq 3$.

Acknowledgements

Éric Marchand's research is supported in part by the Natural Sciences and Engineering Research Council of Canada, and William Strawderman's research is partially supported by a grant from the Simons Foundation (#418098).

References

[1] Baran, J. & Stepień-Baran, A. (2013). Sequential estimation of a location parameter and powers of a scale parameter from delayed observations. *Statistica Neerlandica*, **67**, 263–280.

[2] Berger, J.O. (1975). Minimax estimation of location vectors for a wide class of distributions. *Annals of Statistics*, **3**, 1318–1328.

[3] Brandwein, A.C., Ralescu, S. & Strawderman, W.E. (1993). Shrinkage estimation of the location parameters for certain spherically symmetric distributions. *Annals of the Institute of Statistical Mathematics*, **45**, 551–565.

[4] Brandwein, A.C. & Strawderman, W.E. (1991) Generalizations of James-Stein estimators under spherical symmetry. *Annals of Statistics*, **19**, 1639–1650.

[5] Brandwein, A.C. & Strawderman, W.E. (1980) Minimax estimation of location parameters for spherically symmetric distributions with concave loss. *Annals of Statistics*, **8**, 279–284.

[6] Chung, Y., Kim, C. & Dey, D.K. (1999). A new class of minimax estimators of multivariate normal mean vectors under balanced loss function. *Statistics & Decisions*, **17**, 255–266.

[7] Dey, D., Ghosh, M. & Strawderman, W.E. (1999). On estimation with balanced loss functions. *Statistics & Probability Letters*, **45**, 97–101.

[8] Feller, W. (1966). *An Introduction to Probability Theory and its Applications, volume II.* Second edition, Wiley & Sons, New York.

[9] Fourdrinier, D., Strawderman, W.E. & Wells, M. T. (2018). *Shrinkage estimation.* Springer series in statistics. Springer. New York, Dordrecht, Heidelberg, London.

[10] Hu, G. & Peng, P. (2011). All admissible estimators for a regression coefficient under a balanced loss function. *Journal of Multivariate Analysis*, **102**, 1217–1224

[11] Jafari Jozani, M., Leblanc, A. & Marchand, É (2014). On continuous distribution functions, minimax and best invariant estimators, and integrated balanced loss functions. *Canadian Journal of Statistics*, **42**, 470–486.

[12] Jafari Jozani, M., Marchand, É. & Parsian, A. (2012). Bayesian and robust bayesian analysis under a general class of balanced loss functions. *Statistical Papers*, **53**, 51–60.

[13] Jafari Jozani, M., Marchand, É. & Parsian, A. (2006). On estimation with weighted balanced-type loss function. *Statistics & Probability Letters*, **76**, 773–780.

[14] Judge, G.G. & Mittlehammer, R.C. (2004). A semiparametric basis for combining estimation problems under quadratic loss. *Journal of the American Statistical Association*, **99**, 479–487.

[15] Kubokawa, T., Marchand, É & Strawderman, W.E. (2015). On improved shrinkage estimators under concave loss. *Statistics & Probability Letters*, **96**, 241–246.

[16] Stein, C. (1981). Estimation of the mean of multivariate normal distribution. *Annals of Statistics*, **9**, 1135–1151.

[17] Strawderman, W.E. (2003). On minimax estimation of a normal mean vector for general quadratic loss. *Mathematical Statistics and Applications: Festschrift for Constance van Eeden*, IMS Lecture Notes, 4–14.

[18] Strawderman, W.E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *Journal of Multivariate Analysis*, **8**, 255–264

[19] Zellner, A. (1994). Bayesian and Non-Bayesian estimation using balanced loss functions. *Statistical Decision Theory and Methods V*, (J.O. Berger and S.S. Gupta Eds). New York: Springer-Verlag, 337–390.

[20] Zhang, Q. & Chen, P. (2018). Credibility estimators with dependence structure over risks and time under balanced loss function. *Statistica Neerlandica*, **72**, 157–173.

[21] Zinodiny, S., Rezaeia, S. & Nadarajah, S. (2014). Bayes minimax estimation of the multivariate normal mean vector under balanced loss function. Statistics & Probability Letters, **93**, 96–101.